图书情报工作
第 62 卷 第 17 期 2018 年 9 月



# 科研领域重点研究方向判别方法研究*

邓启平[1] 王小梅[2]

[1] 电子科技大学图书馆 成都 610054 [2] 中国科学院科技战略咨询研究院 北京 100190

**摘要：**[目的/意义]判别重点研究方向对科研管理和科技政策的制定有着重要参考价值，已有的定量方法多是根据创新性、新颖性以及增长性等特征属性设计指标进行识别、推荐，本研究进一步利用研究方向间的关联关系，从网络拓扑结构和特征属性两个维度判别重点研究方向。[方法/过程]在构建领域文献引用网络的基础上，利用大规模网络聚类算法识别研究方向，并构建研究方向关联网络，利用网络重要节点识别算法从网络拓扑结构的角度判别重点研究方向，同时结合新颖性、增长性和 H 指数三个特征属性指标，构建了重点研究方向遴选指标体系。[结果/结论]对纳米科技领域进行实证分析，经专家判读，认为加权 PageRank、Gefura 以及增长性指标更加具有客观性、全面性和稳定性，通过综合运用三个指标遴选出 208 个纳米科技领域的重点研究方向。

**关键词：**重点研究方向 科学结构 网络重要节点 纳米领域

**分类号：**G250

**DOI：**10.13266/j.issn.0252-3116.2018.17.008



## 1 引言

引领未来发展、对科学和技术发展有很强带动作用的科研领域重点研究方向受到国家相关研究发展计划的大力支持，因此，如何识别科研领域中的重点研究方向为科研管理和科技政策的制定提供支撑有着重大意义。重点研究方向是为了解决国家战略需求中的重大科学问题和对人类认识世界将会起到重要作用的科学前沿问题而提出，其在科学计量学界并没有明确的定义，与之相关的是研究前沿及其类似概念（新兴主题、热点研究等），即由科研领域文献聚类得到的一组文献所表征的并具有新生、快速发展特性的研究方向。显然，重点研究方向不应该仅涵盖具有新生和快速发展特性的研究方向。研究方向及其关联关系形成了一个复杂的研究方向关联网络，可以利用网络重要节点识别算法从网络拓扑结构的角度测度研究方向在科研领域中的影响力，由此得到的高影响力研究方向与研究前沿一样也属于重点研究方向的范畴。因此，本研究以研究方向的关联关系为切入点，从网络拓扑结构的角度提出一种新的重点研究方向判别方法，以期对已有方法进行补充，进而更加全面准确地识别出重点研究方向。

## 2 相关研究

早期判别重点研究方向主要是通过定性的方法，随着定量方法的发展，以文献计量为核心通过引文聚类分析识别领域研究方向的新途径，为人们认识科学结构、判别重点研究方向提供了新方法，主要包括研究方向识别和重点研究方向遴选两个部分。研究方向识别是指通过引文聚类分析，从文献数据集中识别领域所包含的研究方向及其之间关联关系的过程，这是判别重点研究方向的前提。重点研究方向的遴选多基于研究方向的特征属性，利用时间线分析研究方向的演化情况，识别新生研究方向，或设计统计指标遴选新颖性和增长性等特征显著的研究方向。实际应用中，引文聚类分析包括共被引、文献耦合和直接引用三种。

### 2.1 基于共被引分析判别重点研究方向

如何基于共被引分析判别重点研究方向，学者已经开展了一系列研究。H. Small 等首先利用单链接聚类算法从高被引论文的共被引关系中识别研究方向，然后将连续时间窗口下的研究方向（高被引论文簇）连接，研究每个方向的演化情况，并利用近期文献数量、平均出版年等指标遴选热点研究方向[1-3]。随后，

* 本文系国家自然科学基金面上项目“科学结构特征及其演化动力学分析方法与应用研究”（项目编号：71173211）研究成果之一。

作者简介：邓启平（ORCID：0000-0001-7078-2026），助理馆员，硕士；王小梅（ORCID：0000-0002-9895-1511），研究馆员，硕士生导师，通讯作者，E-mail：wangxm@casisd.cn。

收稿日期：2018-03-13 修回日期：2018-05-09 本文起止页码：60-67 本文责任编辑：杜杏叶

邓启平，王小梅.科研领域重点研究方向判别方法研究[J].图书情报工作,2018,62(17):60-67.

大量研究借鉴该模型分析重点研究方向，并通过提升文献覆盖范围和聚类分析准确度细分领域结构，进而得到更加全面的分析结果。S. Upham 等从 ISI 收录的 1999-2004 年高被引论文中识别研究前沿，将其分为吸收型和增长型两类，并分析了二者在被引频次和论文增长速度等方面的特征属性[4]。陈超美等人利用文献共被分析描绘领域结构，并借助词频探测算法遴选新兴主题，经过多个领域的实证分析，发现新兴主题存在被引频次快速增长且介数中心性较高的关键文献[5-7]。K. Boyack 等绘制了 2000-2010 年度全学科领域微观层面科学结构(定位到研究问题)，重点分析了石墨烯和染料敏化太阳能电池两个新兴领域的发展趋势及特征[8]。王小梅等开展了科学结构地图系列研究，借鉴 H. Small 的方法分析全学科领域的科学结构特征及其演化规律，并利用篇均被引频次、论文年均增长率等指标遴选热点研究前沿[9-10]。中国科学院科技战略咨询研究院、科睿唯安公司与中国科学院文献情报中心联合发布的“研究前沿报告”[11]，也是将共被引聚类分析得到的高被引论文簇作为研究前沿，并利用被引频次、核心论文数量、论文平均出版年等指标遴选热点前沿、新兴前沿以及重点前沿。

### 2.2 基于耦合或直接引用分析判别重点研究方向

由于相对成熟的方法体系，共被引分析成为了判别重点研究方向的主要方法，但其存在明显的时间滞后性，近年有学者利用文献耦合和直接引用分析判别重点研究方向。S. Morris 在 2003 年以文献耦合聚类分析得到的文献簇代表研究前沿，并以炭疽病研究为例，采用创新性的时间线方式分析领域的发展状况[12]。J. Hopcroft 对 Citeseer 数据的文献耦合网络做聚类分析，通过对比两个时期(1990-1998 年和 1990-2001 年)的聚类结果识别新兴主题[13]。M. Huang 等利用移动时间窗分析方法识别 OLED 领域 2000-2009 年的研究前沿，并根据发展速度将其分为新兴前沿、生长前沿以及稳定前沿等[14]。采用直接引用分析判别重点研究方向可追溯到 E. Garfield 的历史图谱，该方法能够展现研究方向的演化路径，成为了历史编纂学的核心技术；E. Garfield 及其同事后来开发的引文编年可视化系统(HistCite)，可基于直接引用网络生成领域知识结构地图，并分析研究内容发生迁移的重点研究领域[15-17]。N. Shibata 等对复杂网络、再生医学等领域的相关文献逐年进行聚类分析，并基于文献的重叠度连接不同时间窗口下的研究方向，用以展示领域结构的演化，同时利用研究方向的核心文献网络拓扑属性和平均出版年等信息遴选新兴前沿[18-19]。

### 2.3 基于多种引用分析判别重点研究方向

此外，还有学者尝试使用混合引用分析方法探测重点研究方向。H. Small 等分别采用共被引和直接引用两种方法探测 Scopus 数据(1996-2010 年)的科学结构，分析发现直接引用模型更容易区分新兴主题，并以直接引用模型为基础，共被引模型作为调整，利用新颖性和增长性遴选新兴主题[20]。W. Glänzel 等将文献划分到两个独立的时间窗口下，通过文献耦合聚类分析得到研究方向，并利用核心文献的直接引用关系确定不同时间窗口下研究方向的演进关系，同时结合文献量的变化情况判别新生和快速发展的研究方向[21]。

除 K. Boyack 和 H. Small 的研究外[8,20]，上述研究在研究方向探测环节采用的方法仅适用于小规模数据集，且描绘的科学结构较为宏观。重点研究方向遴选方面，已有研究主要利用新颖性、增长性以及引用影响力等特征属性测度研究方向的重要性，进而识别新生或快速发展的研究前沿，很少利用研究方向的关联关系，从网络拓扑结构的角度测度研究方向的重要性。本研究尝试构建一套较为完整的重点研究方向判别方法，能够探测大规模数据集下的微观科学结构，并利用网络重要节点识别算法从网络拓扑结构的角度判别研究方向的重要性，同时结合研究方向的特征属性构建遴选指标体系和遴选策略，全面判别重点研究方向。

## 3 方法和数据

以纳米科技领域作为分析对象，利用 S. Arora 等提出的纳米科技模块化检索式[22]，从 Web of Science 平台下载 2009-2014 年发表的类型为 Article 的文献，清洗后共 570 104 篇。首先利用大规模网络聚类算法识别文献直接引用网络中的研究方向，并以特征词和核心文献相结合的方式表征研究方向的中心主题；然后构建研究方向关联网络，揭示出纳米科技领域的科学结构。在此基础上，引入网络重要节点识别算法从网络拓扑结构的角度判别研究方向在科学领域中的重要性，并结合特征属性指标构建二维遴选指标体系。最后经专家判读，选取相对准确有价值的指标遴选重点研究方向，并将结果进行合并，形成重点研究方向推荐集合。

### 3.1 研究方向识别

基于文献间的直接引用关系构建文献引用网络并提取最大连通网络，借鉴 N. Eck 的方法对引用关系标准化[23]，然后采用 L. Waltman 等提出的 SLM[24] 算法进行聚类分析。选用 Modularity Optimizer 工具包(ht-

tp://www. ludowaltman. nl/slm/)执行 SLM 算法,分解参数设定为 80。最大连通网络包含 505 036 个文献节点和 3 116 367 对引用关系,聚类分析得到 4 237 个类,平均文献数量为 135 篇,选取文献量超过(含)20 篇的 3 612个类作为研究方向。

文本分析是解读研究内容的常用方法,H. Small 与 K. Boyack 等在研究中从文献的题名和摘要中抽取特征词表征研究方向的中心主题[8,19]。由于独立的特征词缺少语义关系,有时难以确定研究方向的中心主题,有学者利用核心文献替代特征词[20]。因此,本研究采用二者结合的方式表征研究方向的中心主题,首先选取前 10 个高频作者关键词作为特征词,然后用 LeaderRank(LR)[25]算法计算最大连通网络中每篇文献的得分,从每个研究方向中选取得分排名前 5 的文献作为核心文献。需要说明的是,LeaderRank 是 PageRank 的改进算法,具备收敛速度快、参数少等优势,更适用于大规模网络。

### 3.2 研究方向关联网络构建

为揭示纳米科技领域的科学结构,并从网络拓扑结构的角度遴选重点研究方向,本研究利用文献间的引用关系构建了研究方向关联网络,方法如图 1 所示。将每个研究方向(论文簇)视为独立的节点,如果研究方向 i 引用了研究方向 j 中的文献,则存在节点 i 指向节点 j 的边,其关联强度用如下公式表示:

$$A_{ij} = \sum_{k}^{L(i)} \sum_{g}^{L(j)} c(k,g) \qquad \text{公式(1)}$$

$c(k,g)$ 为二元取值,如果研究方向 i 中的文献 k 引用了研究方向 j 中的文献 g 则为 1,否则为 0;$L(i)$ 表示研究方向 i 包含的文献数量。A 值越大表明研究方向间的关联关系越强,为规避弱关联的干扰,仅保留了 $A>2$ 的边。利用 Gephi 软件对研究方向关联网络可视化,结果见图 2。

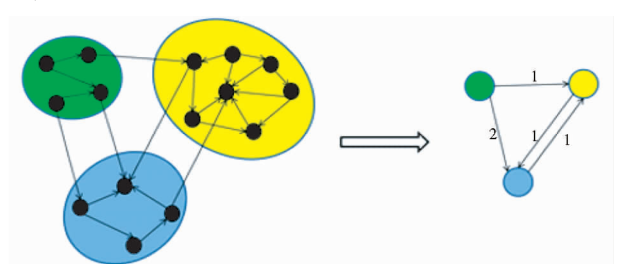



图 1 研究方向关联网络构建示意图

### 3.3 重点研究方向遴选指标设计

在研究方向关联网络中,重要节点识别算法能从网络拓扑结构维度测度研究方向的重要性,这是对基于特征属性遴选重点研究方向的有效补充。显然,仅

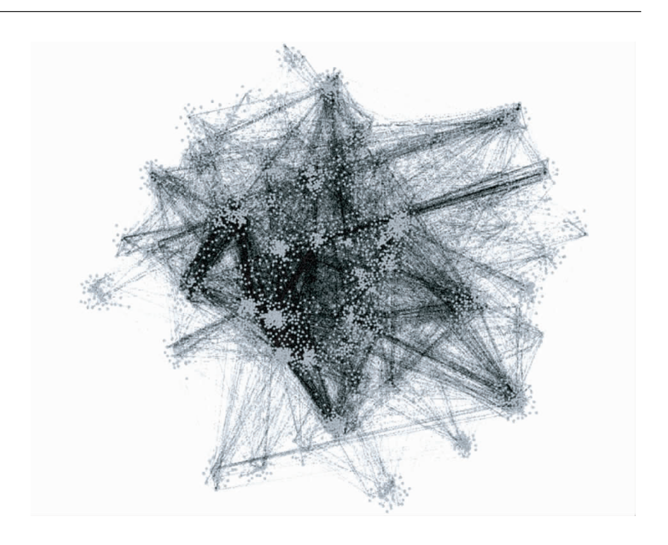


图 2 纳米科技领域研究方向关联网络可视化图谱(基于 OpenOrd 布局算法)

从网络维度遴选重点研究方向具有一定的局限性,有必要进一步结合特征属性,选取两个维度的多个指标遴选重点研究方向,以期更符合科学家的认知。

3.3.1 基于网络拓扑结构遴选重点研究方向 网络中心节点和枢纽节点是两种不同类型的重要节点,在本研究构建的研究方向关联网络中,中心节点是节点度较高的研究方向,其文献被相关研究方向大量引用,具有较高的影响力;枢纽节点是介数中心性较高的研究方向,其与不同的研究方向间存在关联关系并相互影响,通常研究内容会产生重大转变。上述两类节点都是值得关注的重点研究方向,本研究选取相应的重要节点识别算法进行判别。PageRank 算法同时考虑了邻居节点的数量和质量,反映了节点之间相互影响的关系,能更真实地测度研究方向在网络中的影响力;由于本研究构建的是微观科学结构,网络存在显著的组群划分现象,即多个研究方向属于同一研究大类,Gefura 算法是考虑网络内部组群划分现象的变形介数中心性,计算了经过一个节点连接两个不同组群的最短路径数量,值越大说明节点的枢纽作用越强。因此,本研究选取 PageRank 和 Gefura 算法遴选重点研究方向。

(1)PageRank 算法。在含权网络中,PageRank 算法(PR)的计算公式如下[26]:

$$\mathrm{PR}_i(t) = c\sum_{j=1}^{n} w_{ji}\frac{PR_j(t-1)}{\sum_l^n w_{jl}} + \frac{(1-c)}{n} \qquad \text{公式(2)}$$

其中,$w_{ji}$ 表示节点 j 指向节点 i 的权重,n 是网络中节点的个数,c 为阻尼系数,该指标中 c 的取值为 0.85。

(2)Gefura 算法。Gefura 算法的计算公式为[27]:

$$\Gamma_G^B(a) = \frac{1}{M_a}\sum_{\substack{g\in G_k\\ h\in G_l\\ k\neq l}}\left(\frac{p_{g,h}(a)}{p_{g,h}}\right) \qquad \text{公式(3)}$$

其中，$p_{g,h}$ 表示从节点 $g$ 到节点 $h$ 的最短路径数量，$p_{g,h}(a)$ 表示从节点 $g$ 到节点 $h$ 的最短路径中经过节点 $a$ 的数量，$G_k$ 代表组群 $k$ 中的所有节点，$M_a = \sum_{k,l} |G_k \setminus \{a\}| \cdot |G_l \setminus \{a\}|$，代表除去节点 $a$ 后组群 $k$ 和 $l$ 之间可能存在的节点对。为确认每个研究方向所属的组群，利用 *SLM* 算法实现多层级聚类分析，以分解参数 0.8 下的聚类结果作为上层研究大类，根据文献的分布情况将当前的研究方向划分到相应文献占比最高的研究大类中。

3.3.2 基于特征属性遴选重点研究方向　传统的基于属性特征遴选重点研究方向有一些成熟的方法值得借鉴，本研究选取考察了新颖性、增长性和引用影响力三个属性特征的遴选结果。

（1）新颖度指标。新颖度（*DN*）主要是通过论文的平均出版年来表征，研究方向包含的文献的平均出版年距离当前年份越近，越可能是前沿研究。利用标准化后的平均出版年测度新颖度，计算公式如下：

$$APY_i = \frac{1}{N_i} \sum_{j=1}^{N_i} PY_j \qquad \text{公式(4)}$$

$$DN_i = \frac{APY_i - \min(APY)}{\max(APY) - \min(APY)} \qquad \text{公式(5)}$$

$N_i$ 表示研究方向 i 的论文数量，$PY_j$ 为论文 j 的发表年份，$DN_i$ 的取值范围在 0 到 1 之间。

（2）增长度指标。增长度（DG）表征的是研究方向的发展速度，多以文献、主题词、作者数量等的变化情况来测度，本研究以文献年均增长率测度研究方向的增长度，其计算公式为：

$$DG_i = \frac{1}{L_i} * \frac{V_i^c - V_i^f}{V_i^f} \qquad \text{公式(6)}$$

$L_i$ 是研究方向 i 中文献的年份间隔周期，由当前年减去最早出现的年份得到；$V_i^c$ 为研究方向当前年发表的文献数量；$V_i^f$ 为研究方向 i 第一年的发文量。

（3）H 指数。测度研究方向的引用影响力有多种方式，本文选用 H 指数从论文的质量和数量两方面衡量研究方向的引用影响力。研究方向 H 指数的计算方法是将所有论文按被引频次降序排列，满足第 n 篇论文的被引次数恰好不小于 n 的最大 n 值，即为该研究方向的 H 指数，其计算公式如下[28]：

$$C_{n+1} \leqslant n+1, C_n \geqslant n \qquad \text{公式(7)}$$

其中，$C_{n+1}$ 为研究方向内按被引频次降序排列的第 n+1 篇论文的被引频次。H 指数表征了研究方向内的高被引论文数量。

3.3.3 基于综合指标遴选重点研究方向　PR 算法与引用次数密切相关，因此早期发文较多的研究方向 PR 值偏高，而新兴研究方向的 PR 值偏低，为弥补这一不足，通过加权的方式将新颖度指标融合到 PR 算法中进行调和，在网络拓扑结构相似的情况下，新颖度高的研究方向将会拥有更高的得分。原始 PR 算法中，随机游走是以相同的概率（1/n）访问每个节点，分析认为新颖度高的研究方向（近期发文较多）应有更高的概率被访问，因而将随机游走概率设定为与新颖度的大小成正例。加权 PageRank（PR_W）的公式为：

$$PR_W_i(t) = c \sum_{j=1}^{n} w_{ji} \frac{PR_W_j(t-1)}{\sum_l^n w_{jl}} + (1-c) \frac{DN_i}{\sum_l^n DN_l} \qquad \text{公式(8)}$$

$DN_i$ 表示节点 i 的新颖度，阻尼系数 c 设定为0.15。

## 4　结果分析

为了选择合适的指标遴选重点研究方向，本研究首先分析了各个指标的相关性，找出不同指标遴选重点方向结果的差异性，同时结合领域专家意见确定遴选策略，进而遴选纳米科技领域的重点研究方向。

### 4.1　指标相关性分析

利用斯皮尔曼等级相关对 6 个指标的排序结果作相关性分析，结果如表 1 所示。可以看出，PageRank、Gefura、H 指数和加权 PageRank 等 4 个指标的排序结果具有较高的相关性，主要原因是它们都与研究方向的引用相关；新颖度与增长度指标的排序结果高度相关，但二者与其他 4 个指标的相关性较低，具体来看，与 PageRank 和 Gefura 排序的相关性最低，略高于 0.3，其次是 H 指数（略高于 0.4），与加权 PageRank 排序的相关性最高，但也仅达到 0.6。这表明，网络拓扑结构和特征属性两个维度的指标的排序结果存在显著的差异性，特征属性维度的指标，尤其是新颖度和增长度指标，重在遴选快速增长的研究方向，而网络拓扑结构维度的指标聚焦在遴选出早期发文较多且产生了广泛的引用影响力的研究方向；此外，虽然新颖度和加权 PageRank 指标的排序结果相关性不高，但与 PageRank 相比，从 0.330 上升到了 0.648，说明通过新颖度加权的方式对 PageRank 的排序结果起到了调节作用，加权 PageRank 能够同时反映研究方向的影响力和新颖度。

每个指标下排名前 20 的研究方向编号如表 2 所示。PageRank 和 H 指数排名前 20 的研究方向中有 10 是相同的，二者的 TOP20 列表差异较大，看似与相关性分析的结论不一致。通过分析研究方向的具体内容，

**表 1　不同维度指标的相关系数**

| | 网络拓扑结构维度 | | 特征属性维度 | | | 综合指标 |
|---|---|---|---|---|---|---|
| | PageRank | Gefura | 新颖度 | 增长度 | H 指数 | 加权 PageRank |
| PageRank | 1 | 0. 873 | 0. 330 | 0. 331 | 0. 868 | 0. 900 |
| Gefura | | 1 | 0. 354 | 0. 348 | 0. 796 | 0. 821 |
| 新颖度 | | | 1 | 0. 942 | 0. 429 | 0. 648 |
| 增长度 | | | | 1 | 0. 419 | 0. 622 |
| H 指数 | | | | | 1 | 0. 853 |
| 加权 PageRank | | | | | | 1 |

发现 H 指数中不相同的 10 个研究方向规模较大，且研究内容不相似；PageRank 中不相同的 10 个研究方向规模偏小，但研究内容均与石墨烯相关。造成上述现象的原因有 2 个：①H 指数是一个相对独立的统计指标，仅与研究方向自身的规模和文献引用次数相关，H 指数高的研究方向规模较大；②石墨烯是一个宽泛的研究领域，涵盖了许多内容相近的研究方向，这些相近的研究方向形成了一个组群，组群内的研究方向存在密集的引用关系，导致它们的 PR 值较高。还值得注意的是，两个指标下不相同的 10 个研究方向在彼此的排名中都比较靠前（前 100 位）。因此，分析认为 PageRank 和 H 指数的 TOP 列表的差异性是合理的，具备一致性，如果将石墨烯相关的研究视为一个大的研究方向，二者的 TOP 列表将更加一致。虽然 PageRank 与 Gefura 的相关系数高于 H 指数，但从排名的 TOP20 列表看，仅有 7 个相同的研究方向，TOP50 中也只有 20 个研究方向是相同的，表明 PageRank 与 Gefura 的 TOP 列表的差异性较 H 指数更大。新颖度和增长度的 TOP20 列表一致性较高，有 16 个相同的研究方向，但两个指标的 TOP20 列表都与其他 4 个指标体现出了差异性，仅有 1 – 5 个相同的研究方向。加权 PageRank 的 TOP20 列表与 PageRank 更加一致，有 14 个相同的研究方向，但是一些新颖度高的研究方向在加权 PageRank 下的排名有了显著提升，如研究方向 2、201 和 205 等，说明加权 PageRank 算法有效地将 PageRank 和新颖度进行了融合，能综合测度研究方向的重要性。

### 4.2　重点研究方向遴选策略

显然，利用单一的指标难以全面地遴选出重点研究方向，需要综合应用两个维度下的多个指标。考虑到加权 PageRank 能从两个维度综合测度一个研究方向的重要性，应是遴选重点研究方向的主要指标；由于 PageRank 和新颖度被结合在加权 PageRank 中，因此二者不作为遴选指标；H 指数与 PageRank 相关性很高，遴选结果差异不大，也可以不选择；Gefura 和增长度两个指标从不同的角度反映了研究方向的重要性，遴选结果体现出了一定的差异性，应作为遴选指标之一。

**表 2　各指标排名 TOP20 研究方向**

| 排序 | PageRank | Gefura | 新颖度 | 增长度 | H 指数 | 加权 PageRank |
|---|---|---|---|---|---|---|
| 1 | 15 | 144 | 2 | 130 | 0 | 0 |
| 2 | 23 | 0 | 433 | 55 | 2 | 30 |
| 3 | 31 | 31 | 496 | 2 | 9 | 23 |
| 4 | 0 | 153 | 121 | 433 | 10 | 31 |
| 5 | 10 | 9 | 554 | 201 | 8 | 10 |
| 6 | 239 | 164 | 55 | 484 | 5 | 15 |
| 7 | 30 | 3 | 484 | 203 | 7 | 74 |
| 8 | 12 | 238 | 339 | 754 | 23 | 2 |
| 9 | 74 | 54 | 205 | 554 | 29 | 81 |
| 10 | 276 | 130 | 436 | 210 | 15 | 12 |
| 11 | 38 | 20 | 1 122 | 31 | 6 | 8 |
| 12 | 37 | 12 | 9 | 9 | 31 | 38 |
| 13 | 34 | 29 | 201 | 205 | 12 | 239 |
| 14 | 81 | 23 | 203 | 496 | 74 | 174 |
| 15 | 65 | 4 | 111 | 590 | 34 | 153 |
| 16 | 217 | 17 | 31 | 121 | 37 | 183 |
| 17 | 126 | 183 | 130 | 1 122 | 22 | 201 |
| 18 | 8 | 74 | 128 | 13 | 67 | 65 |
| 19 | 140 | 38 | 8 | 111 | 17 | 34 |
| 20 | 46 | 81 | 1 005 | 339 | 130 | 205 |

注：研究方向编号是按规模（论文数量）由大到小的顺序赋值，规模最大的研究方向编号为 0

为了证实上述推断的合理性，邀请国家纳米科学中心的专家对每个指标的排序结果进行判读，专家反馈的意见认为：“在设计的 6 个遴选指标中，增长度、Gefura 和加权 PageRank 更加具有客观性、全面性和稳定性，比较适合对纳米科技领域重点研究方向的判别，建议在以后的研究实践中加以综合运用”。

综上分析，本研究综合运用加权 PageRank、Gefura 和增长度 3 个指标遴选重点研究方向。其中加权 PageRank 作为主要遴选指标，Gefura 作为次要遴选指标，因 Gefura 遴选的是枢纽研究方向，这类研究方向在科

学领域中占比较少。由于前 2 个指标的 TOP 列表基本包含了增长度指标下排名靠前的研究方向,因此将增长度作为辅助遴选指标,用以定位快速发展的重点研究方向,具体的遴选策略包含以下 4 个步骤:

(1)按加权 PageRank 降序排列,选取前 5% 的研究方向作为候补重点研究方向集合 S1。

(2)按 Gefura 降序排列,选取前 3% 的研究方向作为候补重点研究方向集合 S2。

(3)将集合 S1 和集合 S2 进行合并,得到最终的重点研究方向集合 S3。

(4)按增长度指标对集合 S3 降序排列,定位出快速发展的重点研究方向。

### 4.3 纳米科技领域重点研究方向

由于纳米科技领域规模较大,包含的研究方向较多,共遴选出重点研究方向 208 个,经专家判定认为结果具有一定的可靠性。受专业背景知识和篇幅限制,本研究仅选取太阳能电池领域排名前 10 的重点研究方向进行解读。

利用关键词 photovoltaic、solar cell、solar-cell、PV 从清洗后的纳米科技领域文献集合中检索与太阳能电池相关的文献,如果一个研究方向中太阳能电池相关文献占比超过 70%,则认为是太阳能电池领域的研究方向。排名前 10 的太阳能电池重点研究方向如表 3 所示:

表 3　太阳能电池领域 TOP10 重点研究方向

| 研究方向 | 特征词 | 核心文献题名 |
|---|---|---|
| 钙钛矿型太阳能电池 | dye sensitized solar cells, perovskite, solar cells, TiO2, photovoltaic, graft copolymer, perovskite solar cells, solid-state dye-sensitized solar cells, self-assembly, titanium dioxide | Highly efficient nanoporous TiO2-polythiophene hybrid solar cells based on interfacial modification using a metal-free organic dye; Organometal halide perovskites as visible-light sensitizers for photovoltaic cells; An organic D-pi-A dye for record efficiency solid-state sensitized heterojunction solar cells; Sequential deposition as a route to high-performance perovskite-sensitized solar cells; All-solid-state dye-sensitized solar cells with high efficiency |
| 铜锌锡硒/铜锌锡硫薄膜太阳能电池 | solar energy materials, thin films, solar cells, Cu2ZnSnS4, CZTS, Cu2ZnSnSe4, kesterite, optical properties, semiconductors | Synthesis of Cu2ZnSnS4 nanocrystal ink and its use for solar cells; Synthesis of Cu2ZnSnS4 nanocrystals for use in low-cost photovoltaics; Solution-based synthesis and characterization of Cu2ZnSnS4 nanocrystals; Growth and raman scattering characterization of Cu2ZnSnS4 thin films; Fabrication of 7.2% efficient CZTSSe solar cells using CZTS nanocrystals |
| 碳材料做辅助电极的染料敏化太阳能电池 | counter electrode, dye-sensitized solar cell, polyaniline, graphene, platinum, PEDOT, polypyrrole, electrochemistry, DSSC, Electropolymerization | Efficient dye-sensitized cells with catalytic multiwall carbon nanotube counter electrodes; Application of mesoporous carbon to counter electrode for dye-sensitized solar cells; A high-performance counter electrode based on poly(3,4-alkylenedioxythiophene) for dye-sensitized solar cells; Functionalized graphene as a catalytic counter electrode in dye-sensitized solar cells; Carbon nanotubes with titanium nitride as a low-cost counter-electrode material for dye-sensitized solar cells |
| 改进染料敏化太阳能电池中纳米多孔二氧化钛层研究 | dye-sensitized solar cells, light scattering, solar cells, scattering layer, electron transport, TiO2, titania, nanoparticles, nanostructures, photoanode, semiconductors | Mesoporous anatase TiO2 beads with high surface areas and controllable pore sizes: a superior candidate for high-performance dye-sensitized solar cells; TiO2-coated multilayered SnO2 hollow microspheres for dye-sensitized solar cells; Formation of highly efficient dye-sensitized solar cells by hierarchical pore generation with nanoporous TiO2 spheres; Dye-sensitized solar cells employing a single film of mesoporous TiO2 beads achieve power conversion efficiencies over 10%; Dual-function scattering layer of submicrometer-sized mesoporous TiO2 beads for high-efficiency dye-sensitized solar cells |
| 反结构聚合物太阳能电池 | polymer solar cells, organic solar cells, inverted structure, organic photovoltaics, solar cells, graphene oxide, ZnO, stability, MoO3, fullerenes, TiO2 | Enhanced power-conversion efficiency in polymer solar cells using an inverted device structure; Interface engineering for organic electronics; Inverted polymer solar cells integrated with a low-temperature-annealed sol-gel-derived ZnO film as an electron transport layer; Role of tungsten oxide in inverted polymer solar cells; Highly efficient and stable inverted polymer solar cells integrated with a cross-linked fullerene material as an interlayer |
| 表面等离子体增强的有机太阳能电池 | organic solar cells, surface plasmon, silver nanoparticles, light trapping, polymer solar cells, solar cells, plasmonics, organic photovoltaics, bulk heterojunction, nanoparticles, P3HT:PCBM, FDTD | High efficiency polymer solar cells with wet deposited plasmonic gold nanodots; Plasmon-enhanced charge carrier generation in organic photovoltaic films using silver nanoprisms; Surface plasmonic effects of metallic nanoparticles on the performance of polymer bulk heterojunction solar cells; Photonic crystal geometry for organic solar cells; Efficiency enhancement of organic solar cells using transparent plasmonic Ag nanowire electrodes |
| 量子点敏化的太阳能电池 | quantum dots, solar cells, counter electrode, cadmium sulfide, quantum dot-sensitized solar cell, CdS, ZnO, TiO2, silar, photoelectronchemistry | Highly efficient quantum-dot-sensitized solar cell based on co-sensitization of CdS/CdSe; Photosensitization of TiO2 nanostructures with CdS quantum dots: particulate versus tubular support architectures; CdSe quantum dot-sensitized TiO2 electrodes: effect of quantum dot coverage and mode of attachment; PbS and US quantum dot-sensitized solid-state solar cells: "old concepts, new results"; Quantum dot sensitized solar cells. a tale of two semiconductor nanocrystals: CdSe and CdTe |

（续表 3）

| 研究方向 | 特征词 | 核心文献题名 |
| --- | --- | --- |
| 基于二氧化钛纳米棒的染料敏化太阳能电池 | dye-sensitized solar cells, titanium dioxide, solar cells, TiO2, nanorods, hydrothermal synthesis, nanostructure, optical properties, TiO2 nanorods, x-ray diffraction, rutile | Growth of oriented single-crystalline rutile TiO2 nanorods on transparent conducting substrates for dye-sensitized solar cells; Dye-sensitized TiO2 nanotube solar cells with markedly enhanced performance via rational surface engineering; Solar cells by design: photoelectronchemistry of TiO2 nanorod arrays decorated with CdSe; High efficiency dye-sensitized solar cells based on hierarchically structured nanotubes; Enhanced electron collection efficiency in dye-sensitized solar cells based on nanostructured TiO2 hollow fibers |
| 表面等离子体增强太阳能电池光吸收研究 | light trapping, solar cells, surface plasmons, silicon, plasmonics, photovoltaics, nanoparticles, silver nanoparticles, elemental semiconductors, thin film solar cells | Tunable light trapping for solar cells using localized surface plasmons; Design of plasmonic thin-film solar cells with broadband absorption enhancements; Surface plasmon enhancement of optical absorption in thin-film silicon solar cells; Enhancement of optical absorption in thin-film solar cells through the excitation of higher-order nanoparticle plasmon modes; Design considerations for plasmonic photovoltaics |
| 基于氧化锌纳米线的染料敏化太阳能电池 | dye-sensitized solar cells, zinc oxide, solar cells, ZnO, nanoparticles, photoanode, chemical synthesis, semiconductors, nanostructures, ZnO nanorods | Electron transport in dye-sensitized solar cells based on ZnO nanotubes: evidence for highly efficient charge collection and exceptionally rapid dynamics; Nanoforest of hydrothermally grown hierarchical ZnO nanowires for a high efficiency dye-sensitized solar cell; Electrochemical impedance spectroscopy characterization of electron transport and recombination in ZnO nanorod dye-sensitized solar cells; Ordered TiO2 nanotube arrays on transparent conductive oxide for dye-sensitized solar cells; Preferential growth of long ZnO nanowire array and its application in dye-sensitized solar cells |



## 5 总结

本研究采用大规模网络聚类算法识别文献引用网络中的研究方向，并在构建研究方向关联网络的基础上设计遴选指标体系识别重点研究方向。已有的遴选方法主要利用新颖性、增长性以及引用影响力等特征属性识别新生或快速发展的研究方向，本研究提出利用研究方向的关联关系，从网络拓扑结构的角度测度研究方向在科学领域中的重要性，将重要节点识别算法引入研究方向关联网络，选取 PageRank 和 Gefura 两个指标遴选重点研究方向，同时结合研究方向的特征属性，设计新颖度、增长度和 H 指数 3 个指标，并将 PageRank 和新颖度融合提出加权 PageRank 指标，从而构建了二维遴选指标体系。对纳米科技领域实证分析，通过指标相关性分析发现，利用网络重要节点识别算法和特征属性指标遴选的重点研究方向有显著差异，说明两类指标是从不同角度反映研究方向的重要性。专家的判读结果证明了从网络拓扑结构的角度遴选重点研究方向的可行性和有效性，其可以作为已有方法的有效补充。最后本研究结合专家的建议，选取加权 PageRank、Gefura 以及增长度三个指标从纳米科技领域中遴选出 208 个重点研究方向。

与其他相似的研究不同，本研究提出了一种新的重点研究方向判别方法，将方法视角扩展到了网络拓扑结构，是对基于特征属性的判别方法的有效补充。本研究并未对利用重要节点识别算法和特征属性指标得到的不同重点研究方向从本质上加以分析和区分，这是研究中存在的不足之处；此外，研究中仍有待改进的地方，接下来需要优化方法中涉及的参数和重点研究方向的遴选策略，未来可将该方法应用到其他研究领域，从大规模文献集合中判别出重点研究方向。

**参考文献：**


[1] SMALL H. A co-citation model of a scientific specialty: a longitudinal study of collagen research[J]. Social studies science, 1977, 7(2): 139-166.

[2] SMALL H, SWEENEY E. Clustering the science citation index? using co-citations. I. A comparison of methods[J]. Scientometrics, 1985, 7(3-6): 391-409.

[3] SMALL H, SWEENEY E, GREENLEE E. Clustering the science citation index using co-citations. II. Mapping science[J]. Scientometrics, 1985, 8(5/6): 321-340.

[4] UPHAM S, SMALL H. Emerging research fronts in science and technology: patterns of new knowledge development[J]. Scientometrics, 2010, 83(1): 15-38.

[5] CHEN C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.

[6] CHEN C, CHEN Y, HOROWITZ M, et al. Towards an explanatory and computational theory of scientific discovery[J]. Journal of informetrics, 2009, 3(3): 191-209.

[7] CHEN C, HU Z, LIU S, et al. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace[J]. Expert opinion on biological therapy, 2012, 12(5): 593-608.

[8] BOYACK K, KLAVANS R, SMALL H, et al. Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science[J]. Journal of engineering & technology management, 2014, 32(32): 147-159.

[9] 潘教峰，张晓林，王小梅，等. 科学结构地图 2012[M]. 北京：科学出版社，2013.

[10] 王小梅，韩涛，王俊，等. 科学结构地图 2015[M]. 北京：科学出版社，2015.

[11] 中国科学院科技战略咨询研究院. 2017 研究前沿[EB/OL]. [2018－03－12]. http://www.casisd.cn/zkcg/zxcg/201711/P020171121342129165876.pdf.
[12] MORRIS S, YEN G, WU Z, et al. Time line visualization of research fronts[J]. Journal of the American Society for Information Science and Technology, 2003, 54(5): 413－422.
[13] HOPCROFT J, KHAN O, KULIS B, et al. Tracking evolving communities in large linked networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(S1): 5249－5253.
[14] HUANG M, CHANG C. Detecting research fronts in OLED field using bibliographic coupling with sliding window[J]. Scientometrics, 2014, 98(3): 1721－1744.
[15] GARFIELD E. Citation indexes in sociological and historical research[J]. Journal of the American Society for Information Science and Technology, 1963, 14(4): 289－291.
[16] GARFIELD E, PUDOVKIN A, ISTOMIN V. Why do we need algorithmic historiography? [J]. Journal of the American Society for Information Science and Technology, 2003, 54(5): 400－412.
[17] GARFIELD E. Historiographic mapping of knowledge domains literature[J]. Journal of information science, 2004, 30(2): 119－145.
[18] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Detecting emerging research fronts based on topological measures in citation networks of scientific publications[J]. Technovation, 2008, 28(11): 758－775.
[19] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications[J]. Technological forecasting & social change, 2009, 78(2): 274－282.
[20] SMALL H, BOYACK K, KLAVANS R. Identifying emerging topics in science and technology[J]. Research policy, 2014, 43(8): 1450－1467.
[21] GLÄNZEL W, THIJS B. Using 'core documents' for detecting and labelling new emerging topics[J]. Scientometrics, 2012, 91(2): 399－416.
[22] ARORA S, PORTER A, YOUTIE J, et al. Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs[J]. Scientometrics, 2013, 95(1): 351－370.
[23] ECK N, WALTMAN L. CitNetExplorer: a new software tool for analyzing and visualizing citation networks[J]. Journal of informetrics, 2014, 8(4): 802－823.
[24] WALTMAN L, ECK N. A smart local moving algorithm for large-scale modularity-based community detection[J]. European physical journal B, 2013, 86(11): 471.
[25] LÜ L, ZHANG Y, HO Y, et al. Leaders in social networks, the delicious case[J]. Plos one, 2011, 6(6): e21202.
[26] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[J]. Computer networks and ISDN system, 1998, 30(1－7): 107－117.
[27] GUNS R, ROUSSEAU R. Unnormalized and normalized forms of gefura measures in directed and undirected networks[J]. Frontiers of information technology & electronic engineering, 2015, 16(4): 311－320.
[28] HIRSCH J. An index to quantify an individual's scientific research output[J]. Proceedings of the national academy of sciences of the United States of America, 2005, 102(46): 16569－16572.

**作者贡献说明:**
邓启平:负责数据采集与方法实现,撰写论文;
王小梅:设计研究思路,修订论文。



## A Method of Identifying Key Research Directions

Deng Qiping[1]　Wang Xiaomei[2]
[1] Library, University of Electronic Science and Technology of China, Chengdu 611731
[2] Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190

**Abstract:** [**Purpose/significance**] The identification of key research directions (KRDs) is significant to the management of scientific research and the formulation of policy. Existing methods of quantitative analysis mainly used indicators based on novelty, growth and other characteristics to identify and recommend KRDs. This paper used the relationships of research directions further and identified KRDs from network topology and characteristic two dimensions. [**Method/process**] On the basis of the building of citation network of papers in a field, a large-scale network clustering algorithm was used to detect research directions, and the topic association network was built. Then, identification algorithms of important nodes in complex networks were used to identify KRDs from the dimension of network topology, and an index system of selecting KRDs was constructed by combining three indicators of characteristic dimension, including novelty index, growth index and H-index. [**Result/conclusion**] After an empirical analysis of the field of nanotechnology, with the experts' interpretation, the paper concluded that growth index, Gefura and weighted PageRank are more objective and stable. Finally, 208 KRDs were identified by synthetically using above three indicators.

**Keywords:** key research directions　science structure　important node　nanotechnology